Cochrane Database of Systematic Reviews

Testosterone supplementation in men with sexual dysfunction (Protocol)

Jung JH, Kim HW, Narayan V, Reddy B, Walsh T, Yu HS, Lim JS, Kim MH, Dahm P

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Protocol]

Testosterone supplementation in men with sexual dysfunction

Jae Hung Jung[1], Hong Wook Kim[2], Vikram Narayan[3], Balaji Reddy[4], Thomas Walsh[5], Ho Song Yu[6], Jung Soo Lim[7], Myung Ha Kim[8], Philipp Dahm[3,9]

[1]Department of Urology, Yonsei University Wonju College of Medicine, Wonju, Korea, South. [2]Department of Urology, Konyang University College of Medicine,, Daejeon, Korea, South. [3]Department of Urology, University of Minnesota, Minneapolis, Minnesota, USA. [4]Department of Urology, Massachusetts General Hospital, Boston, USA. [5]Immunocompromised Host Section, Pediatric Oncology Branch, National Cancer Institute, Bethesda, MD, USA. [6]Department of Urology, Chonnam National University, Gwangju, Korea, South. [7]Division of Endocrinology, Department of Internal Medicine, Yonsei University Wonju College of Medicine, Wonju, Korea, South. [8]Yonsei Wonju Medical Library, Yonsei University Wonju College of Medicine, Wonju, Korea, South. [9]Urology Section, Minneapolis VA Health Care System, Minneapolis, Minnesota, USA

Contact address: Jae Hung Jung, Department of Urology, Yonsei University Wonju College of Medicine, 20 Ilsan-ro, Wonju, Gangwon, 26426, Korea, South. jungx563@gmail.com, geneuro95@yonsei.ac.kr.

ABSTRACT

This is a protocol for a Cochrane Review (Intervention). The objectives are as follows:

To assess the effects of testosterone compared to placebo or other medical treatments in men with sexual dysfunction.

BACKGROUND

Description of the condition

Sexual dysfunction is any physical or psychological problem that prevents an individual or a couple from getting sexual satisfaction (McCabe 2016). Erectile dysfunction is the most prevalent complaint in men with sexual dysfunction, and its incidence increases with age (Feldman 1994; Hatzimouratidis 2010; Lue 2004). The aging process in men is accompanied by a progressive decline in serum testosterone levels. Serum concentration of testosterone declines by an average of 3.1 to 3.5 ng/dL per year in middle-aged and older men. Low testosterone concentrations (value of less than 325 ng/dL) are seen in 19% of men in their 60s, 28% of

men in their 70s, and 49% of men in their 80s (Harman 2001; Kang 2013). Erectile dysfunction is also associated with reduced or borderline circulating androgen levels (Feldman 2002). Male hypogonadism is traditionally defined as the failure of the testes to produce testosterone, because of a local (testicular; primary) or a distant (pituitary/hypothalamus; secondary) deficiency (Corona 2011). A new classification of male hypogonadism was introduced essentially based on age of onset of the hypogonadal symptoms and signs, thus distinguishing very early (in fetal life), early (pre- or peri-pubertally), and late-onset hypogonadism, which presents after puberty with aging (Corona 2011). Testosterone hormone should be prescribed for men with sexual dysfunction and testosterone deficiency (Khera 2016; Lunenfeld 2013; Wang 2009). Given the controversies in the threshold for diagnosing late-onset

hypogonadism as described in the following section (Seftel 2006), we will consider individuals with sexual dysfunction regardless of serum testosterone level as the target population through which to define the disease condition and intervention in this review.

Diagnosis

Late-onset hypogonadism is a clinical syndrome defined as failure to produce physiological levels of testosterone with several symptoms and signs suggestive of androgen deficiency (Khera 2016). Late-onset hypogonadism is a relatively common medical condition affecting the aging male, but is difficult to diagnose since there are no clear-cut differences in relation to the physiological aging process. However, it has been recognised that diagnosis of late-onset hypogonadism should be made only in men with consistent symptoms and signs and unequivocally low serum testosterone levels (Khera 2016; Lunenfeld 2013; Wang 2009).

The main symptoms of late-onset hypogonadism are low libido and decreased nocturnal and morning erections. Other symptoms of late-onset hypogonadism are several vague and non-specific clinical features such as decreased muscle mass and strength, increased body fat, decreased bone mineral density and osteoporosis, and depressed mood (Khera 2016; Lunenfeld 2013; Wang 2009). Questionnaires such as the Aging Male Symptom Score (AMS) and the Androgen Deficiency in Aging Men that have good sensitivity are widely used in clinical practice to evaluate symptoms related to late-onset hypogonadism, but are not recommended for the diagnosis of hypogonadism because of low specificity (Heinemann 1999; Morley 2000).

The biochemical work-up to establish late-onset hypogonadism is the measurement of serum testosterone. A number of methods are available to assess serum testosterone. Most clinical laboratories estimate the serum concentration of total testosterone by immunoassays which are relatively inexpensive and readily available (Diver 2009). There are no consistent and clinically relevant reference ranges, but total testosterone levels below 12 nmol/L are used as the lower limits of normal total testosterone (Khera 2016; Lunenfeld 2013; Wang 2009). However, approximately 50% to 70% of testosterone in the blood is tightly bound to sex hormone-binding globulin (SHBG); 20% to 30% is loosely bound to albumin; 4% is bound to other proteins; and 1% to 3% is the free, non-bound form in men (Diver 2009). The SHBG-bound form is not readily physiologically available, whereas albumin-bound testosterone and free testosterone (bioavailable testosterone) are able to cross cell membranes and bind to androgen receptors (Diver 2009). Recently, measurement of free or bioavailable testosterone has become the best marker of late-onset hypogonadism and may confirm the diagnosis when the serum total testosterone concentration is not diagnostic of late-onset hypogonadism. Equilibrium dialysis is the gold standard for free testosterone measurement, and free testosterone levels below 225 pmol/L are generally accepted as lower limits of free testosterone (Wang 2009). Alterna-

tively, calculated free testosterone by measuring serum SHBG levels together with total testosterone can be used (Diver 2009; Khera 2016; Lunenfeld 2013; Wang 2009). It has been recommended that blood samples for diagnosing late-onset hypogonadism be obtained in the morning, usually from 7:00 to 11:00 AM, because of the diurnal variation of serum testosterone in which values are highest in the early morning (Diver 2009; Khera 2016; Lunenfeld 2013; Wang 2009). Late blood sampling can overestimate the risk of diagnosing late-onset hypogonadism (Khera 2016).

Treatment

In 1849, Adolph Berthold demonstrated a connection between "testicular secretions", male behaviour, and sexual characteristics (Nieschlag 2014). These testicular secretions have been elucidated as testosterone (Nieschlag 2014; Nieschlag 2015). Since exogenous testosterone administration, known as testosterone supplementation therapy, became available for the Klinefelter and the Kallmann syndrome (Nieschlag 2014; Zitzmann 2000), World Health Organization (WHO) consensus has stated that the major goal of exogenous testosterone administration is to restore testosterone levels at as close to physiologic concentrations associated with the patient's age as possible (Nieschlag 1992).

Testosterone supplementation therapy has been recommended in men who have low testosterone levels associated with certain medical conditions such as failure of the testicles to produce testosterone because of genetic problems or exposures from chemotherapy (primary hypogonadism), or problems with the pituitary gland or hypothalamus that control the production of testosterone by the testis (secondary hypogonadism) (Khera 2016; Lunenfeld 2013; Wang 2009). In addition, testosterone supplementation therapy is also recommended in the treatment of erectile dysfunction and problems with libido arising from low testosterone levels in aging men without other endocrinological causes of low testosterone (Khera 2016; Lunenfeld 2013; Wang 2009).

Based on current guidelines by major professional societies in andrology and endocrinology, there is a general agreement that testosterone supplementation therapy may be offered to symptomatic individuals that indicate a diagnosis of late-onset hypogonadism when circulating total testosterone is below 8 nmol/L (Khera 2016; Lunenfeld 2013; Wang 2009). In the case of a total testosterone level above 12 nmol/L, testosterone supplementation therapy should not be recommend . If the serum total testosterone level is between 8 and 12 nmol/L with typical hypogonadal symptoms, measurement of free testosterone may be helpful to exclude the conditions when SHBG could be decreased (obesity, acromegaly, hypothyroidism) or increased (aging, hepatic illness, hyperthyroidism, use of anticonvulsants).

Description of the intervention

Testosterone has been available for more than 70 years, and the number of new testosterone preparations has increased dramatically over the last two decades (Nieschlag 2014). Recent US data showed that 2.2 million men were prescribed testosterone in 2013 compared to 1.2 million in 2010 (Nguyen 2015). Aggressive marketing strategies by American pharmaceutical companies may explain the remarkable increase in testosterone use, even in the absence of an appropriate diagnosis (Khera 2016; Lunenfeld 2013; Wang 2009).

Currently available testosterone supplementation therapy options include intramuscular injections, transdermal, subcutaneous pellets, oral, and buccal formulations (Giagulli 2011; Shoskes 2016).

- Intramuscular injections is the longest-available therapy option. While relatively inexpensive, frequent administration of testosterone is required to maintain therapeutic levels with two- to three-week intervals owing to their short half-life. Recently, longer-acting testosterone injections have become available with improved compliance compared with the shorter-acting testosterone injections.

- Transdermal formulations (such as gels, creams, and solutions) maintain stable physiological testosterone levels and require daily application. However, there are concerns regarding local adverse events such as skin irritation and rash.

- Subdermal administration (subcutaneous pellet implantation) provides sustained normal testosterone levels for three to four months, but implantation is invasive requiring surgical intervention.

- Other therapies (oral tables and buccal patch) are also available. These formulations are convenient for daily use, but have disadvantages owing to frequent fluctuations in plasma testosterone levels.

Testosterone supplementation therapy can be initiated with any of the approved formulations in accordance with patient preference, predicted adverse effects, cost, and insurance coverage (Shoskes 2016).

Adverse effects of the intervention

The most important concerns about testosterone administration relate to cardiovascular risks in older men with age-related decline in testosterone levels (US Food and Drug Administration). Two observational studies that examined the risk of cardiovascular events associated with testosterone supplementation therapy found an increased risk of cardiovascular harm (Finkle 2014; Vigen 2013). Similarly, a recent meta-analysis of 27 randomised controlled trials reported that testosterone therapy was associated with an increased risk of adverse cardiovascular events (odds ratio 1.5, 95% confidence interval 1.1 to 2.1) (Xu 2013). The US Food and Drug Administration (FDA) has warned that testosterone supplementation therapy increases cardiovascular risk and has cautioned that prescription testosterone products are approved only for men

who have low testosterone levels caused by certain medical conditions (US Food and Drug Administration).

Additonal adverse events of testosterone supplementation therapy are polycythaemia and an increase in prostate-related events. Testosterone supplementation therapy commonly increases serum concentrations of haemoglobin and haematocrit which is described as erythrocytosis or polycythaemia related to the risk for thromboembolism (Jones 2015b). Another health risk of testosterone supplementation therapy is an increased incidence of prostate-related events (i.e. combined incidence of prostate-biopsy, prostate enlargement, elevated prostate specific antigen (PSA), and prostate cancer) (Calof 2005). There are concerns that testosterone supplementation therapy may increase prostate cancer risk, but it is unclear whether there is in fact a causal relation (Coward 2009; Shabsigh 2009).

In addition, exogenous testosterone also produces dose-dependent depression of gonadotropin (follicle-stimulating hormone (FSH) and luteinising hormone (LH)) release either by direct action on the pituitary gland or by suppression of the hypothalamic gonadotropin-releasing hormone (GnRH) release. Reduced gonadotropin secretion results in decreased intratesticular and peripheral testosterone levels, which can cause adverse effects such as testicular atrophy, oligospermia, azoospermia, and other sperm abnormalities (Fronczak 2012). A lack of libido, erectile dysfunction, or even gynecomastia may also be observed (Nieschlag 2015).

How the intervention might work

Physiologic action of testosterone depends on the presence of an intact hypothalamic pituitary gonadal axis. The hypothalamus produces GnRH, which stimulates FSH and LH production in the anterior pituitary gland. Follicle-stimulating hormone is responsible for sperm production, and LH is responsible for testosterone secretion in the testis. Testosterone is secreted primarily by the testes, and to a lesser extent by the adrenal glands. It exerts its action through binding to and activation of the androgen receptor in most organ systems, including the central and peripheral nervous system, musculoskeletal system, reproductive system, and cardiovascular system (Corona 2011).

Testosterone plays a critical role in the development of the male reproductive organs such as the testes and prostate, as well as promoting secondary sexual characteristics during puberty (Gannon 2016; Ohlander 2016). In addition, testosterone modulates male sexual desire, spontaneous sexual thoughts, motivation, attentiveness to erotic stimuli, and sexual activity (Gannon 2016). It also stimulates muscle protein synthesis and may increase bone mineral density by direct effect of testosterone or by an indirect action requiring conversion to estradiol (Borst 2015). Finally, testosterone is thought to exert a variety of beneficial effects on blood vessels and the heart though improvement of lipid profiles, insulin sensitivity, and exercise tolerance (Kloner 2016).

Why it is important to do this review

The benefit and safety of testosterone supplementation therapy for the treatment of men with low testosterone are not well established (Garnick 2015). Testosterone supplementation therapy has been performed in younger androgen deficient men for many years. However, testosterone supplementation therapy for older men with low testosterone levels has long been a topic of discussion (Zitzmann 2000). While there are existing systematic reviews for testosterone supplementation therapy to treat males with sexual dysfunction, none so far have used the same rigorous methodology as Cochrane Reviews, which include the GRADE approach for rating the certainty of evidence (Guo 2016; Tsertsvadze 2009). Given the controversial and important nature of this question, we propose a methodologically rigorous assessment of the benefits and harms of testosterone supplementation therapy in men with sexual dysfunction.

OBJECTIVES

To assess the effects of testosterone compared to placebo or other medical treatments in men with sexual dysfunction.

METHODS

Criteria for considering studies for this review

Types of studies

We will include randomised controlled trials.

Types of participants

We will include adult men (aged 40 years and over) with sexual dysfunction. The age limitation is based on observations from the Massachusetts Male Aging Study, which noted that the prevalence of complete impotence triples from 5% to 15% between 40 and 70 years (Feldman 1994).

Diagnostic criteria for sexual dysfunction

We will define sexual dysfunction as acquired symptoms of decreased libido (or desire to engage in sexual activity) and/or erectile dysfunction referring to inability to attain or maintain an erection sufficient to permit satisfactory sexual performance (Khera 2016; Lunenfeld 2013; Wang 2009). We will include trials that use validated questionnaires related to sexual dysfunction such as AMS or International Index of Erectile Function (IIEF) to establish the diagnosis of sexual dysfunction (Heinemann 1999; Rosen 1997).

Types of interventions

We plan to investigate the following comparisons of intervention versus control/comparator.

Intervention

- Testosterone supplementation therapy regardless of formulation
- Testosterone supplementation therapy + phosphodiesterase 5 inhibitor (PDE5I)

Comparator

- Placebo compared with testosterone supplementation therapy or testosterone supplementation therapy + PDE5I
- PDE5I compared with testosterone supplementation therapy or testosterone supplementation therapy + PDE5I

Concomitant interventions will have to be the same in both the intervention and comparator groups to establish fair comparisons.

Minimum duration of intervention

We define a clinically meaningful minimal duration of the intervention as one month (four weeks).

Minimum duration of follow-up

We define a clinically meaningful minimal duration of follow-up as three months (12 weeks).

Summary of specific exclusion criteria

We will exclude trials of men with the following.
- Primary hypogonadism such as disorders of the testicles.
- Secondary hypogonadism such as disorders of pituitary gland or brain that cause a hypogonadism.
- Disorders that cause hormonal disturbances such as Cushing disease or syndrome.
- Relative or absolute contraindications of testosterone supplementation therapy such as haematocrit > 55%, known prostate cancer or PSA levels greater than 4 ng/mL, and untreated sleep apnoea.

Types of outcome measures

We will not exclude a trial if it fails to report one or several of our primary or secondary outcome measures. If a trial reports none of our primary or secondary outcomes, we will not include the trial but provide some basic information in an additional table.
We will investigate the following outcomes using the methods and time points specified below.

Primary outcomes

- Erectile function
- Sexual quality of life
- Cardiovascular mortality

Secondary outcomes

- Treatment withdrawal due to adverse events
- Prostate-related events
- Lower urinary tract symptoms
- Socioeconomic effects

Method of outcome measurement

We will consider clinically important difference for the review outcomes to rate the certainty of the evidence for imprecision in the 'Summary of findings' table (Johnston 2010). We will use minimal clinically important difference (MCID) for participant-reported outcomes and clinically relevant absolute or relative effect measures for dichotomous outcomes to define clinically important difference.

Erectile function

- Evaluated by validated questionnaires such as sexual domain of AMS, erection domain of IIEF or IIEF-5, or evaluated by IIEF question 15.
- We will consider the MCID in the erectile function domain score of IIEF to be four (Rosen 2011). We will also consider improvement of IIEF-5 to be over five points as MCID (Spaliviero 2010). There is no established threshold in the sexual domain of the AMS. We will use an MCID of 25% improvement from baseline in AMS (Nickel 2015). For IIEF question 15 "In the last 4 weeks, how would you rate your confidence that you get and keep your erection?", we will classify those who answered very low or low as having erectile dysfunction (Wessells 2011). We will use the MCID in the IIEF question 15 as relative risk reduction of at least 25% (Guyatt 2011).

Sexual quality of life

- Evaluated by validated questionnaires such as total score of AMS and overall satisfaction domain of IIEF.
- There is no reported threshold in total score of AMS and overall satisfaction domain of IIEF. We will use an MCID of 25% improvement from baseline to assess sexual quality of life (Nickel 2015).

Cardiovascular mortality

- Such as mortality related to myocardial infarction, heart failure, constrictive pericarditis, stroke, and any type of embolism or thrombosis after participants were randomised to intervention/comparator groups (Basaria 2010).

Treatment withdrawal due to adverse events

- Defined as treatment discontinuation due to adverse events at any time after participants were randomised to intervention/comparator groups.

Prostate-related events

- Such as incidence of prostate-biopsy, prostate enlargement, elevated PSA, and prostate cancer.
- If the authors of eligible studies reported prostate-related events separately, we will use the available information described in the studies.

Lower urinary tract symptoms

- Evaluated by validated questionnaires such as International Prostate Symptom Score (IPSS).
- We will consider the MCID in the IPSS to be three points (Barry 1995).

Socioeconomic effects

- Such as average length of hospital stay due to adverse events (measured in days from admission to discharge) and direct costs defined as visits to general practitioner.

Timing of outcome measurement

We will consider outcomes measured up to and including 12 months after randomisation as short term, and later than 12 months as long term.

Search methods for identification of studies

Electronic searches

We will search the following sources from the inception of each database to the date of search and will place no restrictions on language of publication or publication status.

- Cochrane Central Register of Controlled Trials (CENTRAL) via the Cochrane Register of Studies Online (CRSO)
- MEDLINE Ovid (Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R); from 1946 onwards)
- Embase (via Elsevier) (from 1974 onwards)

- LILACS (Latin American and the Caribbean Health Sciences Literature; www.bireme.br/; from 1982)
- ClinicalTrials.gov (www.clinicaltrials.gov)
- World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) (www.who.int/ trialsearch/)

We will continuously apply a MEDLINE (via Ovid SP) email alert service established by the Cochrane Metabolic and Endocrine Disorders (CMED) group to identify newly published trials using the same search strategy as described for MEDLINE (Appendix 1). After we submit the final review draft for editorial approval, the CMED group will perform a complete search update on all databases available at the editorial office and will send the results to the review authors. Should we identify new trials for inclusion, we will evaluate these and incorporate the findings into our review draft (Beller 2013).

Searching other resources

We will attempt to identify other potentially eligible trials or ancillary publications by searching the reference lists of included trials, systematic reviews, meta-analyses, and health technology assessment reports. In addition, we will contact authors of included trials to identify any additional information on the retrieved trials or any further trials we may have missed.

We will not use abstracts or conference proceedings for data extraction unless full data are available from the trial authors because this information source does not fulfil the CONSORT requirements, which consist of "an evidence-based, minimum set of recommendations for reporting randomized trials" (CONSORT; Scherer 2007). We will list key data from abstracts in an appendix. We will present information on abstracts or conference proceedings in the 'Characteristics of studies awaiting classification' table.

Data collection and analysis

Selection of studies

Two review authors (JHJ, HWK, BR, JSL, or HSY) will independently screen the abstract, title, or both of every record retrieved by the literature searches to determine which trials should be assessed further. We will obtain the full text of all potentially relevant records. Any disagreements will be resolved through consensus or by recourse to a third review author (PD). In the event that a disagreement cannot be resolved, we will categorise the trial as a 'study awaiting classification' and will contact the trial authors for clarification. We will present an adapted PRISMA flow diagram to show the process of trial selection (Liberati 2009). We will list all articles excluded after full-text assessment in a 'Characteristics of excluded studies' table and will provide the reasons for their exclusion.

Data extraction and management

Two review authors (JHJ, HWK, BR, JSL, or HSY) will independently extract key participant and intervention characteristics for the included trials. We will describe interventions according to the 'template for intervention description and replication' (TIDieR) checklist (Hoffmann 2014; Hoffmann 2017).

We will report data on efficacy outcomes and adverse events using standardised data extraction sheets from the CMED group. Any disagreements will be resolved by discussion or by consulting a third review author (PD) if necessary.

We will provide information including trial identifier for potentially relevant ongoing trials in the 'Characteristics of ongoing studies' table and in a joint appendix 'Matrix of trial endpoint (publications and trial documents)'. We will attempt to locate the protocol for each included trial and will report primary, secondary, and other outcomes in comparison with data in publications in a joint appendix.

We will email authors of all included trials to ask if they would be willing to answer questions regarding their trials. We will present the results of this survey in an appendix. We will thereafter seek relevant missing information on the trial from the primary trial author(s), if required.

Dealing with duplicate and companion publications

In the event of duplicate publications, companion documents, or multiple reports of a primary trial, we will maximise the information yield by collating all available data and will use the most complete data set aggregated across all known publications. We will list duplicate publications, companion documents, multiple reports of a primary trial, and trial documents of included trials (such as trial registry information) as secondary references under the study ID of the included trial. Furthermore, we will also list duplicate publications, companion documents, multiple reports of a trial, and trial documents of excluded trials (such as trial registry information) as secondary references under the study ID of the excluded trial.

Data from clinical trials registers

If data from included trials are available as study results in clinical trials registers such as ClinicalTrials.gov or similar sources, we will make full use of this information and extract the data. If there is also a full publication of the trial, we will collate and critically appraise all available data. If an included trial is marked as a completed study in a clinical trial register but no additional information (study results, publication, or both) is available, we will add this trial to the 'Characteristics of studies awaiting classification' table.

Assessment of risk of bias in included studies

Two review authors (JHJ, HWK, BR, JSL, or HSY) will independently assess the risk of bias for each included trial. Any disagreements will be resolved by consensus or by consulting a third review author (PD) when necessary. In the case of disagreement, we will consult the remainder of the review author team and make a judgement based on consensus. If adequate information is not available from the publications, trial protocols, or other sources, we will contact the trial authors for more detail to request missing data on 'Risk of bias' items.

We will use the Cochrane 'Risk of bias' assessment tool (Higgins 2017), assigning assessments of low, high or unclear risk of bias (for details see Appendix 2). We will evaluate individual bias items as described in the *Cochrane Handbook for Systematic Reviews of Interventions* according to the criteria and associated categorisations contained therein (Higgins 2017).

Summary assessment of risk of bias

We will present a 'Risk of bias' graph and a 'Risk of bias' summary figure.

We will distinguish between self reported and investigator-assessed and adjudicated outcome measures.

We will consider the following self reported outcomes.
- Erectile function
- Sexual quality of life
- Lower urinary tract symptoms

We will consider the following outcomes to be investigator assessed.
- Cardiovascular mortality
- Treatment withdrawal due to adverse event
- Prostate-related events
- Socioeconomic effects

Risk of bias for a trial across outcomes

Some 'Risk of bias' domains, such as selection bias (sequence generation and allocation sequence concealment), affect the risk of bias across all outcome measures in a trial. In case of high risk of selection bias, we will mark all endpoints investigated in the associated trial as being at high risk. Otherwise, we will not perform a summary assessment of the risk of bias across all outcomes for a trial.

Risk of bias for an outcome within a trial and across domains

We will assess the risk of bias for an outcome measure by including all entries relevant to that outcome (i.e. both trial-level entries and outcome-specific entries). We will consider low risk of bias to denote a low risk of bias for all key domains; unclear risk to denote an unclear risk of bias for one or more key domains; and high risk to denote a high risk of bias for one or more key domains.

Risk of bias for an outcome across trials and across domains

These are the main summary assessments that we will incorporate into our judgements about the certainty of the evidence in the 'Summary of findings' tables. We will define outcomes as at low risk of bias when most information comes from trials at low risk of bias; unclear risk when most information comes from trials at low or unclear risk of bias; and high risk when a sufficient proportion of information comes from trials at high risk of bias.

Measures of treatment effect

When at least two included trials are available for a comparison and a given outcome, we will attempt to express dichotomous data as a risk ratio (RR) or odds ratio (OR) with 95% confidence intervals (CIs). For continuous outcomes measured on the same scale (e.g. weight loss in kg), we will estimate the intervention effect using the mean difference (MD) with 95% CIs. For continuous outcomes that measure the same underlying concept (e.g. sexual quality of life) but use different measurement scales, we will calculate the standardised mean difference (SMD). We will express time-to-event data as a hazard ratio (HR) with 95% CIs.

Unit of analysis issues

We will take into account the level at which randomisation occurred, such as cross-over trials, cluster-randomised trials, and multiple observations for the same outcome. If more than one comparison from the same trial is eligible for inclusion in the same meta-analysis, we will either combine groups to create a single pair-wise comparison or appropriately reduce the sample size so that the same participants do not contribute data to the meta-analysis more than once (splitting the 'shared' group into two or more groups). While the latter approach offers some solution to adjusting the precision of the comparison, it does not account for correlation arising from the same set of participants being in multiple comparisons (Higgins 2011a).

We will attempt to reanalyse cluster-randomised trials that have not appropriately adjusted for potential clustering of participants within clusters in their analyses. The variance of the intervention effects will be inflated by a design effect. Calculation of a design effect involves estimation of an intracluster correlation coefficient (ICC). We will obtain estimates of ICCs by contacting trial authors or imputing the ICC values by using either estimates from other included trials that report ICCs or external estimates from empirical research (e.g. Bell 2013). We plan to examine the impact of clustering using sensitivity analyses.

Dealing with missing data

If possible, we will obtain missing data from the authors of the included trials. We will carefully evaluate important numerical data such as screened, randomly assigned participants as well as intention-to-treat, and as-treated and per-protocol populations. We

will investigate attrition rates (e.g. dropouts, losses to follow-up, withdrawals) and will critically appraise issues concerning missing data and use of imputation methods (e.g. last observation carried forward).

In trials where the standard deviation (SD) of the outcome is not available at follow-up or we cannot recreate it, we will standardise by the mean of the pooled baseline SD from those trials that reported this information.

Where included trials do not report means and SDs for outcomes and we are unable to obtain the necessary information from trial authors, we will impute these values by estimating the mean and variance from the median, range, and the size of the sample (Hozo 2005).

We will investigate the impact of imputation on meta-analyses by performing sensitivity analyses, and we will report for every outcome which trials had imputed SDs.

Assessment of heterogeneity

In the event of substantial clinical or methodological heterogeneity, we will not report trial results as the pooled effect estimate in a meta-analysis.

We will identify heterogeneity (inconsistency) by visually inspecting the forest plots and by using a standard Chi² test with a significance level of α = 0.1 (Deeks 2017). In view of the low power of this test, we will also consider the I² statistic, which quantifies inconsistency across trials, to assess the impact of heterogeneity on the meta-analysis (Higgins 2002; Higgins 2003).

When we find heterogeneity, we will attempt to determine the possible reasons for it by examining individual trial and subgroup characteristics.

Assessment of reporting biases

If 10 or more trials are included that investigate a particular outcome, we will use funnel plots to assess small-trial effects. Several explanations may account for funnel plot asymmetry, including true heterogeneity of effect with respect to trial size, poor methodological design (and hence bias of small trials), and publication bias (Sterne 2017). We will therefore interpret the results carefully (Sterne 2011).

Data synthesis

We plan to undertake (or display) a meta-analysis only if we judge participants, interventions, comparisons, and outcomes to be sufficiently similar to ensure an answer that is clinically meaningful. Unless good evidence shows homogeneous effects across trials of different methodological quality, we will primarily summarise low risk of bias data using a random-effects model (Wood 2008). We will interpret random-effects meta-analyses with due consideration to the whole distribution of effects and present a prediction interval (Borenstein 2017a; Borenstein 2017b; Higgins 2009). A

prediction interval needs at least three trials to be calculated and specifies a predicted range for the true treatment effect in an individual trial (Riley 2011). For rare events such as event rates below 1%, we will use Peto's odds ratio method, provided there is no substantial imbalance between intervention and comparator group sizes and intervention effects are not exceptionally large. In addition, we will perform statistical analyses according to the statistical guidelines presented in the *Cochrane Handbook for Systematic Reviews of Interventions* (Deeks 2017).

Subgroup analysis and investigation of heterogeneity

We expect the following characteristics to introduce clinical heterogeneity, and plan to carry out the following subgroup analyses limited to the primary outcomes including investigation of interactions (Altman 2003).

- Participant age (less than 65 years versus ≥ 65 years).
- Baseline testosterone serum level (less than 8 nmol/L versus ≥ 8 nmol/L) obtained between 7:00 and 11:00 AM.
- Presence or absence of at least one of the metabolic syndrome components defined as the standard criteria at the time of the beginning of the trial (Grundy 2005).

These subgroup analyses are based on the following observations.

- Sexual dysfunction is strongly associated with age (Feldman 1994; O'Leary 2003). Tolerability of testosterone in elderly men remains controversial, but may differ by age as testosterone treatment might be associated with increased cardiovascular risks (Basaria 2010; Snyder 2016). The age cut-off is based on the WHO definition of old age (WHO 2002).
- In clinical practice, testosterone supplementation therapy can be administered to people with borderline and low testosterone levels, but appears most meaningful in the latter group of individuals (Khera 2016; Lunenfeld 2013; Wang 2009).
- The prevalence of erectile dysfunction was positively associated with the metabolic syndrome, and serum testosterone levels were also significantly lower in men with the metabolic syndrome (Garcia-Cruz 2013). The magnitude of the effect of testosterone supplementation therapy may therefore differ in men with metabolic syndrome (Isidori 2005).

Sensitivity analysis

We plan to perform sensitivity analyses to explore the influence of the following factors (when applicable) on effect sizes by restricting analysis to the following.

- Published trials.
- Effect of risk of bias, as specified in the Assessment of risk of bias in included studies section.
- Very long or large trials, in order to establish the extent to which they dominate the results.

- Using the following filters: diagnostic criteria, imputation, language of publication, source of funding (industry versus other), or country.

We will also test the robustness of results by repeating the analyses using different measures of effect size (RR, OR, etc.) and different statistical models (fixed-effect and random-effects models).

Certainty of the evidence

We will present the overall certainty of the evidence for each outcome specified below according to the GRADE approach, which takes into account issues related not only to internal validity (risk of bias, inconsistency, imprecision, publication bias) but also to external validity, such as directness of results. Two review authors (JHJ, HWK, BR, JSL, or HSY) will independently rate the certainty of the evidence for each outcome. Any differences in assessment will be resolved by discussion or by consulting a third review author (PD).

We will include an appendix entitled 'Checklist to aid consistency and reproducibility of GRADE assessments' to help with standardisation of the 'Summary of findings' tables (Meader 2014). Alternatively, we will use GRADEpro Guideline Development Tool (GDT) software and will present evidence profile tables as an appendix (GRADEpro GDT 2015). We will present results for the outcomes as described in the Types of outcome measures section. If meta-analysis is not possible, we will present the results in a narrative format in the 'Summary of findings' table. We will justify all decisions to downgrade the quality of trials using footnotes, and will make comments to aid the reader's understanding of the Cochrane Review where necessary.

'Summary of findings' table

We will present a summary of the evidence in a 'Summary of findings' table. This will provide key information about the best estimate of the magnitude of the effect, in relative terms and as absolute differences, for each relevant comparison of alternative management strategies, numbers of participants and trials addressing each important outcome, and a rating of overall confidence in effect estimates for each outcome. We will create the 'Summary of findings' table based on the methods described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Schünemann 2017), employing Review Manager 5 table editor (RevMan 2014). We will report the following outcomes, listed according to priority.

1. Erectile function
2. Sexual quality of life
3. Cardiovascular mortality
4. Treatment withdrawal for any reason
5. Prostate-related events
6. Lower urinary tract symptoms
7. Socioeconomic effects

ACKNOWLEDGEMENTS

REFERENCES

Additional references

Altman 2003
Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* 2003;**326**(7382):219.

Barry 1995
Barry MJ, Williford WO, Chang Y, Machi M, Jones KM, Walker-Corkery E, et al. Benign prostatic hyperplasia specific health status measures in clinical research: how much change in the American Urological Association symptom index and the benign prostatic hyperplasia impact index is perceptible to patients?. *Journal of Urology* 1995;**154**(5):1770–4.

Basaria 2010
Basaria S, Coviello AD, Travison TG, Storer TW, Farwell WR, Jette AM, et al. Adverse events associated with testosterone administration. *New England Journal of Medicine* 2010;**363**(2):109–22.

Bell 2013
Bell ML, McKenzie JE. Designing psycho-oncology randomised trials and cluster randomised trials: variance components and intra-cluster correlation of commonly used psychosocial measures. *Psycho-oncology* 2013;**22**:1738–47.

Beller 2013
Beller EM, Chen JK, Wang UL, Glasziou PP. Are systematic reviews up-to-date at the time of publication?. *Systematic Reviews* 2013;**2**:36.

Borenstein 2017a
Borenstein M, Higgins JP, Hedges LV, Rothstein HR. Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods* 2017;**8**(1):5–18.

Borenstein 2017b
Borenstein M. Prediction intervals. www.meta-analysis.com/prediction (accessed prior to 3 July 2017).

Borst 2015
Borst SE, Yarrow JF. Injection of testosterone may be safer

and more effective than transdermal administration for combating loss of muscle and bone in older men. *American Journal of Physiology. Endocrinology and Metabolism.* 2015; **308**(12):E1035–42.

Boutron 2014

Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. *Journal of Clinical Oncology* 2014;**32**(36):4120–6.

Calof 2005

Calof OM, Singh AB, Lee ML, Kenny AM, Urban RJ, Tenover JL, et al. Adverse events associated with testosterone replacement in middle-aged and older men: a meta-analysis of randomized, placebo-controlled trials. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 2005;**60**(11):1451–7.

CONSORT

The CONSORT statement. www.consort-statement.org (accessed prior to 22 September 2017).

Corbett 2014

Corbett MS, Higgins JP, Woolacott NF. Assessing baseline imbalance in randomised trials: implications for the Cochrane risk of bias tool. *Research Synthesis Methods* 2014; **5**(1):79–85.

Corona 2011

Corona G, Rastrelli G, Forti G, Maggi M. Update in testosterone therapy for men. *Journal of Sexual Medicine* 2011;**8**(3):639-54; quiz 655.

Coward 2009

Coward RM, Simhan J, Carson CC 3rd. Prostate-specific antigen changes and prostate cancer in hypogonadal men treated with testosterone replacement therapy. *BJU International* 2009;**103**(9):1179–83.

Deeks 2017

Deeks JJ, Higgins JP, Altman DG, editor(s) on behalf of the Cochrane Statistical Methods Group. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JP, Churchill R, Chandler J, Cumpston MS, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (updated June 2017). The Cochrane Collaboration, 2017. Available at www.training.cochrane.org/handbook.

Diver 2009

Diver MJ. Laboratory measurement of testosterone. *Frontiers of Hormone Research* 2009;**37**:21–31.

Feldman 1994

Feldman HA, Goldstein I, Hatzichristou DG, Krane RJ, McKinlay JB. Impotence and its medical and psychosocial correlates: results of the Massachusetts male aging study. *Journal of Urology* 1994;**151**(1):54–61.

Feldman 2002

Feldman HA, Longcope C, Derby CA, Johannes CB, Araujo AB, Coviello AD, et al. Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the Massachusetts male aging study. *Journal of Clinical Endocrinology and Metabolism* 2002;**87**(2):589–98.

Finkle 2014

Finkle WD, Greenland S, Ridgeway GK, Adams JL, Frasco MA, Cook MB, et al. Increased risk of non-fatal myocardial infarction following testosterone therapy prescription in men. *PLOS ONE* 2014;**9**(1):e85805.

Fronczak 2012

Fronczak CM, Kim ED, Barqawi AB. The insults of illicit drug use on male fertility. *Journal of Andrology* 2012;**33**(4): 515–28.

Gannon 2016

Gannon JR, Walsh TJ. Testosterone and sexual function. *Urologic Clinics of North America* 2016;**43**(2):217–22.

Garcia-Cruz 2013

Garcia-Cruz E, Leibar-Tamayo A, Romero J, Piqueras M, Luque P, Cardenosa O, et al. Metabolic syndrome in men with low testosterone levels: relationship with cardiovascular risk factors and comorbidities and with erectile dysfunction. *Journal of Sexual Medicine* 2013;**10**(10):2529–38.

Garnick 2015

Garnick MB. Testosterone replacement therapy faces FDA scrutiny. *JAMA* 2015;**313**(6):563–4.

Giagulli 2011

Giagulli VA, Triggiani V, Corona G, Carbone D, Licchelli B, Tafaro E, et al. Evidence-based medicine update on testosterone replacement therapy (TRT) in male hypogonadism: focus on new formulations. *Current Pharmaceutical Design* 2011;**17**(15):1500–11.

GRADEpro GDT 2015 [Computer program]

McMaster University (developed by Evidence Prime, Inc.). GRADEpro GDT. Version (accessed 12 February 2018). Hamilton (ON): McMaster University (developed by Evidence Prime, Inc.), 2015.

Grundy 2005

Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA, et al. Diagnosis and management of the metabolic syndrome: an American Heart Association/ National Heart, Lung, and Blood Institute scientific statement. *Circulation* 2005;**112**(17):2735–52.

Guo 2016

Guo C, Gu W, Liu M, Peng BO, Yao X, Yang B, et al. Efficacy and safety of testosterone replacement therapy in men with hypogonadism: a meta-analysis study of placebo-controlled trials. *Experimental Therapeutic Medicine* 2016; **11**(3):853–63.

Guyatt 2011

Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence - imprecision. *Journal of Clinical Epidemiology* 2011;**64**:1283–93.

Harman 2001

Harman SM, Metter EJ, Tobin JD, Pearson J, Blackman MR. Longitudinal effects of aging on serum total and free

testosterone levels in healthy men. Baltimore Longitudinal Study of Aging. *Journal of Clinical Endocrinology & Metabolism* 2001;**86**(2):724–31.

Hatzimouratidis 2010
Hatzimouratidis K, Amar E, Eardley I, Giuliano F, Hatzichristou D, Montorsi F, et al. Guidelines on male sexual dysfunction: erectile dysfunction and premature ejaculation. *European Urology* 2010;**57**(5):804–14.

Heinemann 1999
Heinemann LAJ, Zimmermann T, Vermeulen A, Thiel C, Hummel W. A new 'aging males' symptoms' rating scale. *Aging Male* 1999;**2**(2):105–14.

Higgins 2002
Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002;**21**(11):1539–58.

Higgins 2003
Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analysis. *BMJ* 2003;**327** (7414):557–60.

Higgins 2009
Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2009; **172**(1):137–59.

Higgins 2011a
Higgins JPT, Deeks JJ, Altman DG. Chapter 16: Special topics in statistics. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Higgins 2017
Higgins JP, Altman DG, Sterne JAC, editor(s). Chapter 8: Assessing risk of bias in included studies. In: Higgins JP, Churchill R, Chandler J, Cumpston MS, editor (s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (updated June 2017). The Cochrane Collaboration, 2017. Available at www.training.cochrane.org/handbook.

Hoffmann 2014
Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;**348**:g1687.

Hoffmann 2017
Hoffmann TC, Oxman AD, Ioannidis JP, Moher D, Lasserson TJ, Tovey DI, et al. Enhancing the usability of systematic reviews by improving the consideration and description of interventions. *BMJ* 2017;**358**:j2998.

Hozo 2005
Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology* 2005;**5**:13.

Hróbjartsson 2013
Hróbjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *Canadian Medical Association Journal* 2013;**185** (4):E201–11.

Isidori 2005
Isidori AM, Giannetta E, Gianfrilli D, Greco EA, Bonifacio V, Aversa A, et al. Effects of testosterone on sexual function in men: results of a meta-analysis. *Clinical Endocrinology* 2005;**63**(4):381–94.

Johnston 2010
Johnston BC, Thorlund K, Schunemann HJ, Xie F, Murad MH, Montori VM, et al. Improving the interpretation of quality of life evidence in meta-analyses: the application of minimal important difference units. *Health and Quality of Life Outcomes* 2010;**8**:116.

Jones 2015a
Jones CW, Keil LG, Holland WC, Caughey MC, Platts-Mills TF. Comparison of registered and published outcomes in randomized controlled trials: a systematic review. *BMC Medicine* 2015;**13**:282.

Jones 2015b
Jones SD Jr, Dukovac T, Sangkum P, Yafi FA, Hellstrom WJ. Erythrocytosis and polycythemia secondary to testosterone replacement therapy in the aging male. *Sexual Medicine Reviews* 2015;**3**(2):101–12.

Kang 2013
Kang S, Park HJ, Park NC. Serum total testosterone level and identification of late-onset hypogonadism: a community-based study. *Korean Journal of Urology* 2013;**54** (9):619–23.

Khera 2016
Khera M, Adaikan G, Buvat J, Carrier S, El-Meliegy A, Hatzimouratidis K, et al. Diagnosis and treatment of testosterone deficiency: recommendations from the Fourth International Consultation on Sexual Medicine (ICSM 2015). *Journal of Sexual Medicine* 2016;**13**(12):1787–804.

Kirkham 2010
Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;**340**:c365.

Kloner 2016
Kloner RA, Carson C 3rd, Dobs A, Kopecky S, Mohler ER 3rd. Testosterone and cardiovascular disease. *Journal of the American College of Cardiology* 2016;**67**(5):545–57.

Liberati 2009
Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine* 2009;**6**(7):1–28.

Lue 2004

Lue TF, Giuliano F, Montorsi F, Rosen RC, Andersson KE, Althof S, et al. Summary of the recommendations on sexual dysfunctions in men. *Journal of Sexual Medicine* 2004;**1**(1): 6–23.

Lunenfeld 2013

Lunenfeld B, Mskhalaya G, Kalinchenko S, Tishova Y. Recommendations on the diagnosis, treatment and monitoring of late-onset hypogonadism in men - a suggested update. *Aging Male* 2013;**16**(4):143–50.

Mathieu 2009

Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009;**302**:977–84.

McCabe 2016

McCabe MP, Sharlip ID, Atalla E, Balon R, Fisher AD, Laumann E, et al. Definitions of sexual dysfunctions in women and men: a consensus statement from the Fourth International Consultation on Sexual Medicine 2015. *Journal of Sexual Medicine* 2016;**13**(2):135–43.

Meader 2014

Meader N, King K, Llewellyn A, Norman G, Brown J, Rodgers M, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic Reviews* 2014;**3**:82.

Morley 2000

Morley JE, Charlton E, Patrick P, Kaiser FE, Cadeau P, McCready D, et al. Validation of a screening questionnaire for androgen deficiency in aging males. *Metabolism* 2000; **49**(9):1239–42.

Nguyen 2015

Nguyen CP, Hirsch MS, Moeny D, Kaul S, Mohamoud M, Joffe HV. Testosterone and "age-related hypogonadism" - FDA concerns. *New England Journal of Medicine* 2015;**373** (8):689–91.

Nickel 2015

Nickel JC, Brock GB, Herschorn S, Dickson R, Henneges C, Viktrup L. Proportion of tadalafil-treated patients with clinically meaningful improvement in lower urinary tract symptoms associated with benign prostatic hyperplasia - integrated data from 1,499 study participants. *BJU International* 2015;**115**(5):815–21.

Nieschlag 1992

Nieschlag E, Wang C, Handelsman DJ, Swerdloff RS, Wu FC, Einer-Jensen N, et al. Guidelines for the use of androgens. Special programme of research, development and research training in human reproduction of the World Health Organization. World Health Organization, Geneva 1992.

Nieschlag 2014

Nieschlag E, Nieschlag S. Testosterone deficiency: a historical perspective. *Asian Journal of Andrology* 2014;**16** (2):161–8.

Nieschlag 2015

Nieschlag E, Vorona E. Mechanisms in endocrinology: medical consequences of doping with anabolic androgenic steroids: effects on reproductive functions. *European Journal of Endocrinology* 2015;**173**(2):R47–58.

O'Leary 2003

O'Leary MP, Rhodes T, Girman CJ, Jacobson DJ, Roberts RO, Lieber MM, et al. Distribution of the Brief Male Sexual Inventory in community men. *International Journal of Impotence Research* 2003;**15**(3):185–91.

Ohlander 2016

Ohlander SJ, Lindgren MC, Lipshultz LI. Testosterone and male infertility. *Urologic Clinics of North America* 2016;**43** (2):195–202.

RevMan 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Riley 2011

Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;**342**:d549.

Rosen 1997

Rosen RC, Riley A, Wagner G, Osterloh IH, Kirkpatrick J, Mishra A. The International Index of Erectile Function (IIEF): a multidimensional scale for assessment of erectile dysfunction. *Urology* 1997;**49**(6):822–30.

Rosen 2011

Rosen RC, Allen KR, Ni X, Araujo AB. Minimal clinically important differences in the erectile function domain of the International Index of Erectile Function scale. *European Urology* 2011;**60**(5):1010–6.

Scherer 2007

Scherer RW, Langenberg P, von EE. Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews* 2007, Issue 2. DOI: 10.1002/ 14651858.MR000005.pub3

Schünemann 2017

Schünemann HJ, Oxman AD, Higgins JP, Vist GE, Glasziou P, Akl E, et al. on behalf of the Cochrane GRADEing Methods Group and the Cochrane Statistical Methods Group. Chapter 11: Completing 'Summary of findings' tables and grading the confidence in or quality of the evidence. In: Higgins JP, Churchill R, Chandler J, Cumpston MS, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (updated June 2017). The Cochrane Collaboration, 2017. Available at www.training.cochrane.org/handbook.

Seftel 2006

Seftel A. Male hypogonadism. Part II: etiology, pathophysiology, and diagnosis. *International Journal of Impotence Research* 2006;**18**(3):223–8.

Shabsigh 2009

Shabsigh R, Crawford ED, Nehra A, Slawin KM. Testosterone therapy in hypogonadal men and potential prostate cancer risk: a systematic review. *International Journal of Impotence Research* 2009;**21**(1):9–23.

Shoskes 2016

Shoskes JJ, Wilson MK, Spinner ML. Pharmacology of testosterone replacement therapy preparations. *Translational Andrology and Urology* 2016;**5**(6):834–43.

Snyder 2016

Snyder PJ, Bhasin S, Cunningham GR, Matsumoto AM, Stephens-Shields AJ, Cauley JA, et al. Effects of testosterone treatment in older men. *New England Journal of Medicine* 2016;**374**(7):611–24.

Spaliviero 2010

Spaliviero M, Strom KH, Gu X, Araki M, Culkin DJ, Wong C. Does Greenlight HPS laser photoselective vaporization prostatectomy affect sexual function?. *Journal of Endourology* 2010;**24**(12):2051–7.

Sterne 2011

Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;**343**:d4002.

Sterne 2017

Sterne JAC, Egger M, Moher D, Boutron I, editor (s). Chapter 10: Addressing reporting biases. In: Higgins JP, Churchill R, Chandler J, Cumpston MS, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (updated June 2017). The Cochrane Collaboration, 2017. Available at www.training.cochrane.org/handbook.

Tsertsvadze 2009

Tsertsvadze A, Fink HA, Yazdi F, MacDonald R, Bella AJ, Ansari MT, et al. Oral phosphodiesterase-5 inhibitors and hormonal treatments for erectile dysfunction: a systematic review and meta-analysis. *Annals of Internal Medicine* 2009; **151**(9):650–61.

US Food and Drug Administration

US Food, Drug Administration (FDA). FDA Drug Safety Communication: FDA cautions about using testosterone products for low testosterone due to aging; requires labeling change to inform of possible increased risk of heart attack and stroke with use. www.fda.gov/Drugs/DrugSafety/ucm436259.htm (accessed prior to 22 September 2017).

Vigen 2013

Vigen R, O'Donnell CI, Baron AE, Grunwald GK, Maddox TM, Bradley SM, et al. Association of testosterone therapy with mortality, myocardial infarction, and stroke in men with low testosterone levels. *JAMA* 2013;**310**(17):1829–36.

Wang 2009

Wang C, Nieschlag E, Swerdloff R, Behre HM, Hellstrom WJ, Gooren LJ, et al. ISA, ISSAM, EAU, EAA and ASA recommendations: investigation, treatment and monitoring of late-onset hypogonadism in males. *International Journal of Impotence Research* 2009;**21**(1):1–8.

Wessells 2011

Wessells H, Penson DF, Cleary P, Rutledge BN, Lachin JM, McVary KT, et al. Effect of intensive glycemic therapy on erectile function in men with type 1 diabetes. *Journal of Urology* 2011;**185**(5):1828–34.

WHO 2002

World Health Organization. Definition of an older or elderly person. www.who.int/healthinfo/survey/ageingdefnolder/en/ (accessed prior to 22 September 2017).

Wood 2008

Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;**336** (7644):601–5.

Xu 2013

Xu L, Freeman G, Cowling BJ, Schooling CM. Testosterone therapy and cardiovascular events among men: a systematic review and meta-analysis of placebo-controlled randomized trials. *BMC Medicine* 2013;**11**:108.

Zitzmann 2000

Zitzmann M, Nieschlag E. Hormone substitution in male hypogonadism. *Molecular and Cellular Endocrinology* 2000; **161**(1-2):73–88.

* *Indicates the major publication for the study*

APPENDICES

Appendix 1. Search strategies

Cochrane Library (via Wiley) search strategy

#1. MeSH descriptor: [Erectile Dysfunction] explode all trees
#2. Erectile dysfunction* or Erectile function* or Impoten* or Penile erect* or ED:ti,ab,kw (Word variations have been searched)
#3. ((sex or sexual*) near/3 (function* or dysfuc* or problem* or symptom* or arous*)):ti,ab,kw (Word variations have been searched)
#4. #1 or #2 or #3
#5. MeSH descriptor: [Testosterone] explode all trees
#6. "testosterone enanthate" or "testosterone undecanoate" or "testosterone Propionate" or Androgel or Axiron or Fortesta or Testim or Nebido or Andriol:ti,ab,kw (Word variations have been searched)
#7. MeSH descriptor: [Androgens] explode all trees
#8. Androderm or Testoderm or Methyltestosterone or Natesto or Striant:ti,ab,kw (Word variations have been searched)
#9. #5 or #6 or #7 or #8
#10. #4 and #9

MEDLINE (Ovid) search strategy

#1. exp Erectile Dysfunction/
#2. Erectile dysfunction$.tw.
#3. Erectile function$.tw.
#4. Impoten$.tw.
#5. Penile erect$.tw.
#6. ED.tw.
#7. ((sex or sexual$) adj3 (function$ or dysfuc$ or problem$ or symptom$ or arous$)).tw
#8. 1 or 2 or 3 or 4 or 5 or 6 or 7
#9. exp Testosterone/
#10. testosterone enanthate.tw.
#11. testosterone undecanoate.tw.
#12. testosterone Propionate.tw.
#13. Androgel.tw.
#14. Axiron.tw.
#15. Fortesta.tw.
#16. Testim.tw.
#17. Nebido.tw.
#18. Andriol.tw.
#19. exp Androgens/
#20. Androderm.tw.
#21. Testoderm.tw.
#22. Methyltestosterone.tw.
#23. Natesto.tw.
#24. Striant.tw.
#25. 9 -24 OR
#26. 8 and 25
#27. randomized controlled trial.pt.
#28. controlled clinical trial.pt.
#29. randomized.ab.

(Continued)

#30. placebo.ab.
#31. drug therapy.fs.
#32. randomly.ab.
#33. trial.ab.
#34. groups.ab.
#35. 27 - 34 OR
#36. exp animals/ not humans.sh.
#37. 30 not 31
#38. 26 and 37

Embase (via Elsevier) search strategy

#1. 'erectile dysfunction'/exp
#2. 'Erectile dysfunction*':ab,ti
#3. 'Erectile function*':ab,ti
#4. Impoten*:ab,ti
#5. 'Penile erect*':ab,ti
#6. ED:ab,ti
#7. ((sex or sexual*) NEAR/3 (function* or dysfuc* or problem* or symptom* or arous*)):ab,ti
#8. #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7
#9. 'testosterone'/exp
#10. 'testosterone enanthate':ab,ti
#11. 'testosterone undecanoate':ab,ti
#12. 'testosterone Propionate':ab,ti
#13. Androgel:ab,ti
#14. Axiron:ab,ti
#15. Fortesta:ab,ti
#16. Testim:ab,ti
#17. Nebido:ab,ti
#18. Andriol:ab,ti
#19. 'androgen'/exp
#20.. Androderm:ab,ti
#21. Testoderm:ab,ti
#22. Methyltestosterone:ab,ti
#23. Natesto:ab,ti
#24. Striant:ab,ti
#25. #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23
OR #24
#26. #8 AND #25
#27. crossover procedure':de OR 'double-blind procedure':de OR 'randomized controlled trial':de OR 'single-blind procedure':de
OR random*:de,ab,ti OR factorial*:de,ab,ti OR crossover*:de,ab,ti OR ((cross NEXT/1 over*):de,ab,ti) OR placebo*:de,ab,ti OR (
(doubl* NEAR/1 blind*):de,ab,ti) OR ((singl* NEAR/1 blind*):de,ab,ti) OR assign*:de,ab,ti OR allocat*:de,ab,ti OR volunteer*:de,
ab,ti
#28. ('animals'/exp) NOT ('humans'/exp and 'animals'/exp)
#29. #27 NOT #28
#30. #26 AND #29

LILAC search strategy

#1. (mh:("Erectile Dysfunction")) OR (tw:("erectile dysfunction" OR "erectile function" OR impotence OR "penile erection" OR ED OR "sexual dysfuction" OR "sexual function" OR "sexual problem" OR "sexual symptom" OR "sexual arousal"))

#2. (mh:(Testosterone OR Androgens)) OR (tw:(testosterone OR "testosterone enanthate" OR "testosterone undecanoate" OR "testosterone Propionate" OR androgel OR axiron OR fortesta OR testim OR nebido OR andriol OR androgens OR androderm OR testoderm OR methyltestosterone OR natesto OR striant))

#3. 1 AND 2

#4. "((PT:""randomized controlled trial"" OR PT:""controlled clinical trial"" OR PT:""multicenter study"" OR MH:""randomized controlled trials as topic"" OR MH:""controlled clinical trials as topic"" OR MH:""multicenter study as topic"" OR MH:""random allocation"" OR MH:""double-blind method"" OR MH:""single-blind method"") OR ((ensaio$ OR ensayo$ OR trial$) AND (azar OR acaso OR placebo OR control$ OR aleat$ OR random$ OR enmascarado$ OR simpleciego OR ((simple$ OR single OR duplo$ OR doble$ OR double$) AND (cego OR ciego OR blind OR mask))) AND clinic$)) AND NOT (MH:animals OR MH:rabbits OR MH:rats OR MH:primates OR MH:dogs OR MH:cats OR MH:swine OR PT:""in vitro"")"

#5. 3 AND 4

ClinicalTrials.gov

#1. ("erectile dysfunction" OR "erectile function" OR impotence OR "penile erection" OR ED OR "sexual dysfuction" OR "sexual function" OR "sexual problem" OR "sexual symptom" OR "sexual arousal")

#2. (testosterone OR "testosterone enanthate" OR "testosterone undecanoate" OR "testosterone Propionate" OR androgel OR axiron OR fORtesta OR testim OR nebido OR andriol OR androgens OR androderm OR testoderm OR methyltestosterone OR natesto OR striant)

#3. 1 AND 2

World Health Organization International Clinical Trials Registry Platform Search Portal

#1. In the title = ("erectile dysfunction" OR "erectile function" OR impotence OR "penile erection" OR ED OR "sexual dysfuction" OR "sexual function" OR "sexual problem" OR "sexual symptom" OR "sexual arousal") AND In the intervention= (testosterone OR "testosterone enanthate" OR "testosterone undecanoate" OR "testosterone Propionate" OR androgel OR axiron OR fORtesta OR testim OR nebido OR andriol OR androgens OR androderm OR testoderm OR methyltestosterone OR natesto OR striant)

Appendix 2. 'Risk of bias' assessment

'Risk of bias' domains

Random sequence generation (selection bias due to inadequate generation of a randomised sequence)
For each included trial, we will describe the method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it should produce comparable groups

- Low risk of bias: the trial authors achieved sequence generation using computer-generated random numbers or a random numbers table. Drawing of lots, tossing a coin, shuffling cards or envelopes, and throwing dice are adequate if an independent person performed this who was not otherwise involved in the trial. We will consider the use of the minimisation technique as equivalent to being random.
- Unclear risk of bias: insufficient information about the sequence generation process.
- High risk of bias: the sequence generation method was non-random or quasi-random (e.g. sequence generated by odd or even date of birth; sequence generated by some rule based on date (or day) of admission; sequence generated by some rule based on

(Continued)

hospital or clinic record number; allocation by judgement of the clinician; allocation by preference of the participant; allocation based on the results of a laboratory test or a series of tests; or allocation by availability of the intervention).

Allocation concealment (selection bias due to inadequate concealment of allocation prior to assignment)

We will describe for each included trial the method used to conceal allocation to interventions prior to assignment and will assess whether intervention allocation could have been foreseen in advance of or during recruitment, or changed after assignment

- Low risk of bias: central allocation (including telephone, interactive voice-recorder, web-based and pharmacy-controlled randomisation); sequentially numbered drug containers of identical appearance; sequentially numbered, opaque, sealed envelopes.
- Unclear risk of bias: insufficient information about the allocation concealment.
- High risk of bias: used an open random allocation schedule (e.g. a list of random numbers); assignment envelopes used without appropriate safeguards; alternation or rotation; date of birth; case record number; any other explicitly unconcealed procedure.

We will also evaluate trial baseline data to incorporate assessment of baseline imbalance into the 'Risk of bias' judgement for selection bias (Corbett 2014). Chance imbalances may also affect judgements on the risk of attrition bias. In the case of unadjusted analyses, we will distinguish between trials rated as at low risk of bias on the basis of both randomisation methods and baseline similarity, and trials judged as being at low risk of bias on the basis of baseline similarity alone (Corbett 2014). We will reclassify judgements of unclear, low, or high risk of selection bias as specified in Appendix 3.

Blinding of participants and study personnel (performance bias due to knowledge of the allocated interventions by participants and personnel during the trial)

We will evaluate the risk of detection bias separately for each outcome (Hróbjartsson 2013). We will note whether endpoints were self reported, investigator-assessed, or adjudicated outcome measures (see below)

- Low risk of bias: blinding of participants and key study personnel was ensured, and it was unlikely that the blinding could have been broken; no blinding or incomplete blinding, but we judge that the outcome is unlikely to have been influenced by lack of blinding.
- Unclear risk of bias: insufficient information about the blinding of participants and study personnel; the trial does not address this outcome.
- High risk of bias: no blinding or incomplete blinding, and the outcome is likely to have been influenced by lack of blinding; blinding of trial participants and key personnel attempted, but it is likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding.

Blinding of outcome assessment (detection bias due to knowledge of the allocated interventions by outcome assessment)

We will evaluate the risk of detection bias separately for each outcome (Hróbjartsson 2013). We will note whether endpoints were self reported, investigator-assessed, or adjudicated outcome measures (see below)

- Low risk of bias: blinding of outcome assessment is ensured, and it is unlikely that the blinding could have been broken; no blinding of outcome assessment, but we judge that the outcome measurement is unlikely to have been influenced by lack of blinding.
- Unclear risk of bias: insufficient information about the blinding of outcome assessors; the trial did not address this outcome.
- High risk of bias: no blinding of outcome assessment, and the outcome measurement was likely to have been influenced by lack of blinding; blinding of outcome assessment, but it is likely that the blinding could have been broken, and the outcome measurement was likely to be influenced by lack of blinding.

Incomplete outcome data (attrition bias due to amount, nature, or handling of incomplete outcome data)

For each included trial or each outcome, or both, we will describe the completeness of data, including attrition and exclusions from the analyses. We will state whether the trial reported attrition and exclusions, and report the number of participants included in the analysis at each stage (compared with the number of randomised participants per intervention/comparator groups). We will also note if the trial reported the reasons for attrition or exclusion and whether missing data were balanced across groups or were related to outcomes. We will consider the implications of missing outcome data per outcome such as high dropout rates (e.g. above 15%) or disparate attrition rates (e.g. difference of 10% or more between trial arms)

- Low risk of bias: no missing outcome data; reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to introduce bias); missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups; for dichotomous outcome data, the proportion of missing outcomes compared with observed event risk was not enough to have a clinically relevant impact on the intervention effect estimate; for continuous outcome data, plausible effect size (mean difference or standardised mean difference) among missing outcomes was not enough to have a clinically relevant impact on observed effect size; appropriate methods, such as multiple imputation, were used to handle missing data.

- Unclear risk of bias: insufficient information to assess whether missing data in combination with the method used to handle missing data were likely to induce bias; the trial did not address this outcome.
- High risk of bias: reason for missing outcome data was likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; for dichotomous outcome data, the proportion of missing outcomes compared with observed event risk was enough to induce clinically relevant bias in the intervention effect estimate; for continuous outcome data, plausible effect size (mean difference or standardised mean difference) among missing outcomes was enough to induce clinically relevant bias in observed effect size; 'as-treated' or similar analysis done with substantial departure of the intervention received from that assigned at randomisation; potentially inappropriate application of simple imputation.

Selective reporting (reporting bias due to selective outcome reporting)

We will assess outcome reporting bias by integrating the results of the appendix 'Matrix of trial endpoints (publications and trial documents)' (Boutron 2014; Jones 2015a; Mathieu 2009), with those of the appendix 'High risk of outcome reporting bias according to the Outcome Reporting Bias In Trials (ORBIT) classification' (Kirkham 2010). This analysis will form the basis for the judgement of selective reporting

- Low risk of bias: the trial protocol was available and all the trial's prespecified (primary and secondary) outcomes that were of interest to this review were reported in the prespecified way; the study protocol was unavailable, but it was clear that the published reports included all expected outcomes (ORBIT classification).
- Unclear risk of bias: insufficient information about selective reporting.
- High risk of bias: not all the trial's prespecified primary outcomes were reported; one or more primary outcomes were reported using measurements, analysis methods, or subsets of the data (e.g. subscales) that were not prespecified; one or more reported primary outcomes were not prespecified (unless clear justification for their reporting was provided, such as an unexpected adverse effect); one or more outcomes of interest in this review were reported incompletely so that we could not enter them in a meta-analysis; the trial report failed to include results for a key outcome that would have been expected to have been reported for such a trial (ORBIT classification).

Other bias

- Low risk of bias: the trial appears to be free from other sources of bias.
- Unclear risk of bias: there was insufficient information to assess whether an important risk of bias existed; insufficient rationale or evidence that an identified problem introduced bias.
- High risk of bias: the trial had a potential source of bias related to the specific trial design used; the trial was claimed to be fraudulent; or the trial had some other serious problem.

Appendix 3. Selection bias decisions

Selection bias decisions for trials that reported unadjusted analyses: comparison of results obtained using method details alone with results using method details and trial baseline information[a]

Reported randomisation and allocation concealment methods	Risk of bias judgement using methods reporting	Information gained from study characteristics data	Risk of bias using baseline information and methods reporting
Unclear methods	Unclear risk	Baseline imbalances present for important prognostic variable(s)	**High risk**
		Groups appear similar at baseline for all important prognostic variables	**Low risk**

		Limited or no baseline details	Unclear risk
Would generate a truly random sample, with robust allocation concealment	Low risk	Baseline imbalances present for important prognostic variable (s)	**Unclear risk**[b]
		Groups appear similar at baseline for all important prognostic variables	Low risk
		Limited baseline details, showing balance in some important prognostic variables[c]	Low risk
		No baseline details	**Unclear risk**
Sequence is not truly randomised, or allocation concealment is inadequate	High risk	Baseline imbalances present for important prognostic variable (s)	High risk
		Groups appear similar at baseline for all important prognostic variables	**Low risk**
		Limited baseline details, showing balance in some important prognostic variables[c]	**Unclear risk**
		No baseline details	High risk

[a]Taken from Corbett 2014; judgements highlighted in bold indicate situations in which the addition of baseline assessments would change the judgement about risk of selection bias, compared with using methods reporting alone.
[b]Imbalance identified that appears likely to be due to chance.
[c]Details for the remaining important prognostic variables are not reported

CONTRIBUTIONS OF AUTHORS

All review authors contributed to, read, and approved the final protocol draft.

DECLARATIONS OF INTEREST

Jae Hung Jung (JHJ): none known.

Hong Wook Kim (HWK): none known.

Vikram Narayan (VN): none known.

Balaji Reddy (BR): none known.

Thomas Walsh (TW): consultancy (Boston Scientific), stock/stock options (Counsyl Genetics).

Ho Song Yu (HSY): none known.

Jung Soo Lim (JSL): none known.

Myung Ha Kim (MHK): none known.

Philipp Dahm (PD): none known.

SOURCES OF SUPPORT

Internal sources

- Department of Urology, Yonsei University Wonju College of Medicine, Korea, South.
- Department of Urology, University of Minnesota, Minneapolis, Minnesota, USA.
- Urology Section, Minneapolis VA Health Care System, Minneapolis, Minnesota, USA.

External sources

- No sources of support supplied

NOTES

We have based parts of the Methods, as well as Appendix 1, Appendix 2, and Appendix 3 of this Cochrane Review protocol on a standard template established by the Cochrane Metabolic and Endocrine Disorders group.